



Explainable AI-Based Robust Prediction of Oocyte Retrieval Using AMH Data in ART Cycles



Hyo-Jin Park¹ · Jaewon Lee² · Kyuman Jeong³ · Sehyun Park³ · Deog-Bon Koo¹

¹ Department of Infertility Medical Industry, Daegu University | ² Medical IT Convergence Research Section, ETRI | ³ Department of AI, Daegu University

69,255
ART cycles screened

27,435
COS cycles analyzed

0.8374
Measured-AMH AUC

35.5%
AUC 0.80 threshold

BACKGROUND & OBJECTIVE

- Anti-Müllerian hormone (AMH) is an important predictor of ovarian response, but it is often unavailable in real-world ART records.
- Complete-case analysis excludes these cycles and may limit the generalizability of prediction models.

STUDY AIM

Determine whether AI-imputed AMH preserves oocyte-retrieval prediction when AMH measurements are missing.

STUDY POPULATION & WORKFLOW

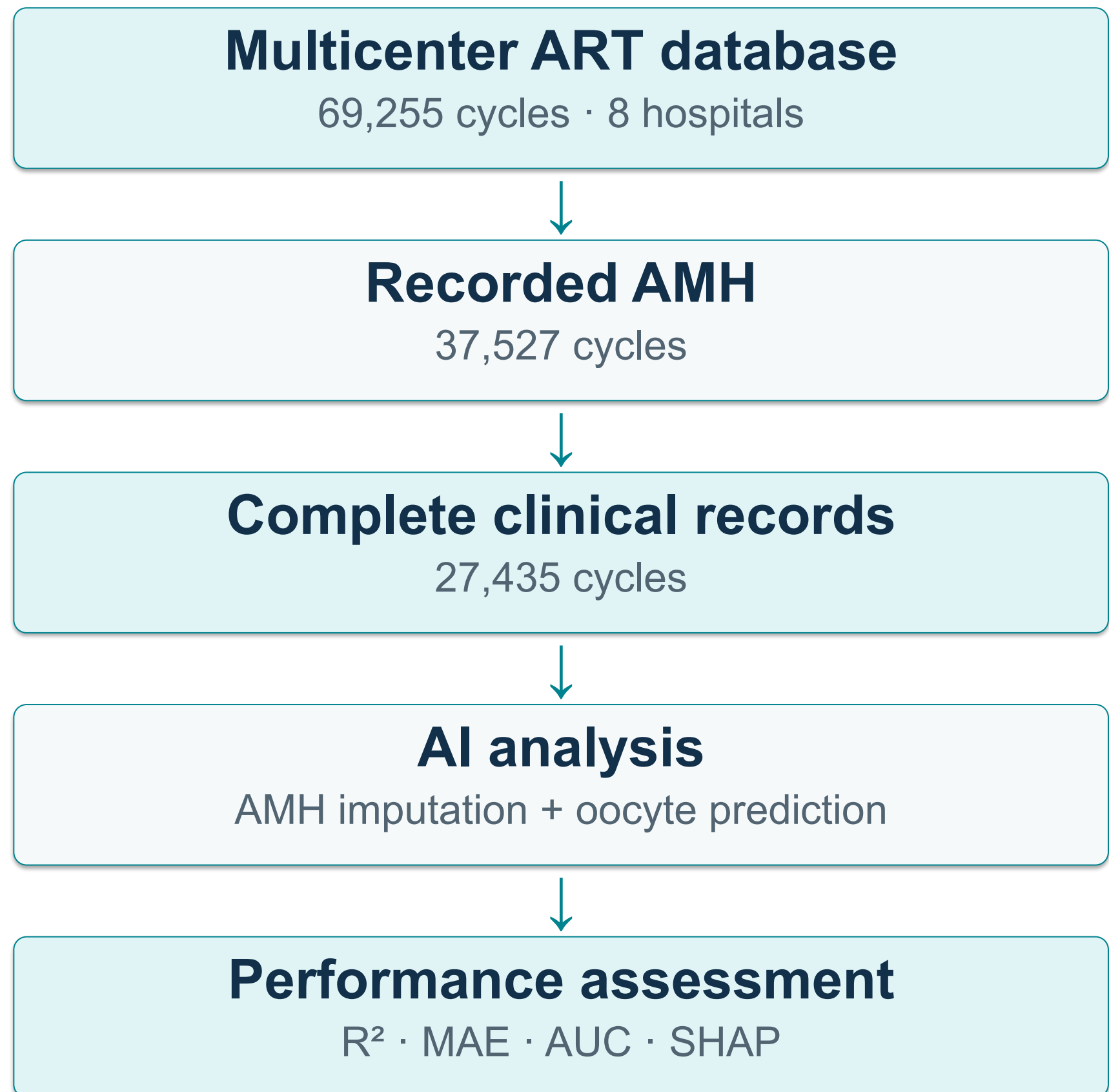


Fig. 1. Cohort selection and analysis workflow. AMH masking was simulated after complete-case selection.

STUDY DESIGN

- Retrospective cohort: ART cycles from 2018–2023
- Inputs: age, AMH, infertility cause, stimulation protocol, FSH and hCG administration
- Four algorithms: Linear regression, XGBoost, LightGBM, and DNN
- Train/validation split: 80% / 20%
- Low response: <8 retrieved oocytes; normal response: ≥8

CLINICAL CONTEXT

The median age was 37 years [34–40], median AMH was 2.41 ng/mL [1.26–4.39], and median retrieved oocyte count was 10 [6–15]. Most analyzed cycles used a GnRH-antagonist protocol.

Primary question

Is accurate reconstruction of AMH itself essential, or is preserving downstream ovarian-response prediction more important?

METHODS: AI MODELS

- Missing AMH values were randomly masked in otherwise complete records at rates from 0% to 100% (10% increments).
- Each model estimated AMH from available variables; imputed AMH was then used to predict retrieved oocytes.
 - Continuous prediction: R² and mean absolute error (MAE)
 - Low versus normal response: area under ROC curve (AUC)
 - Feature contributions: SHAP interpretation

MODEL COMPARISON

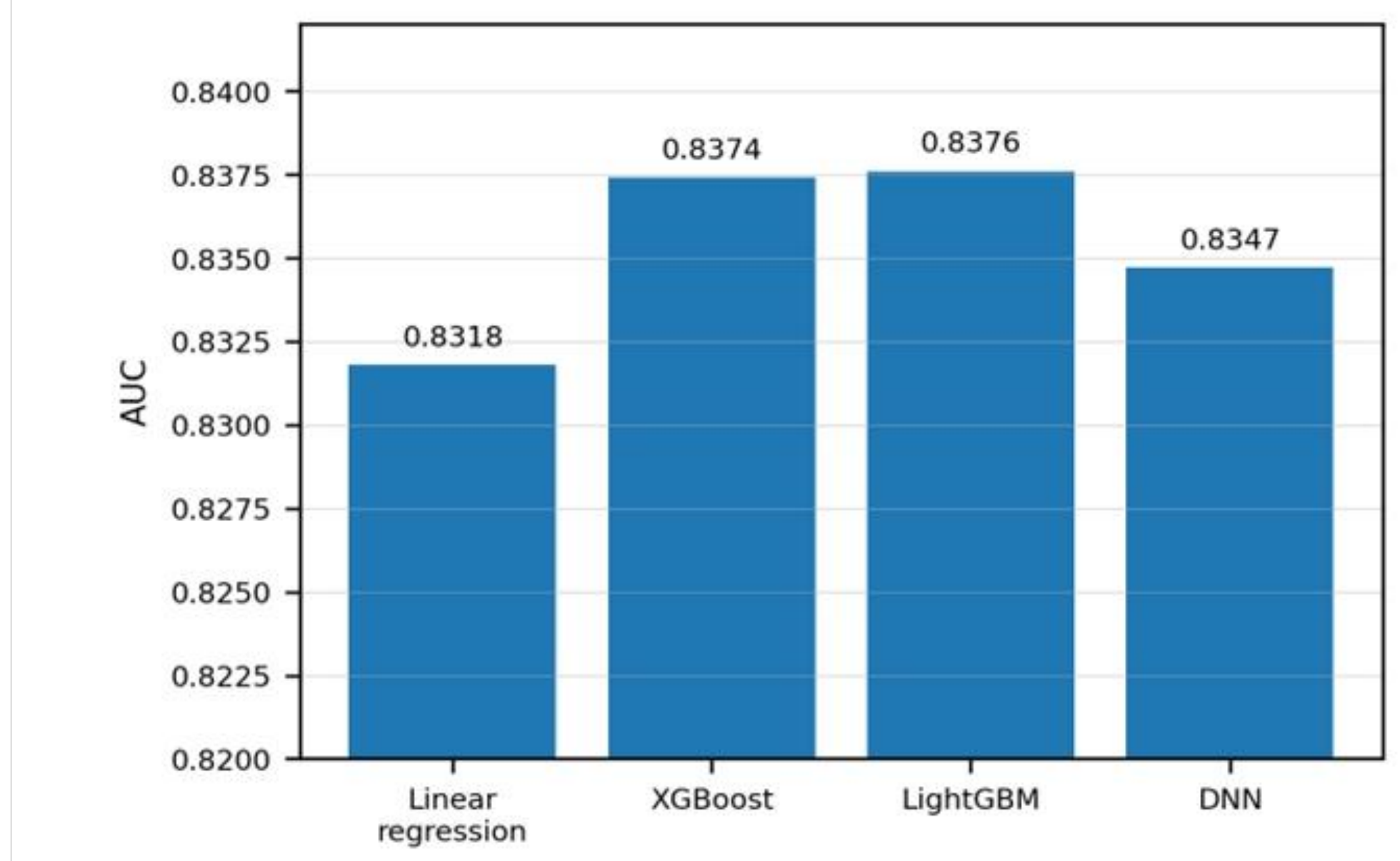


Fig. 2. AUC by model, measured AMH. LightGBM had the highest AUC (0.8376), while XGBoost achieved 0.8374.

XGBoost: R² = 0.4063 | MAE = 3.9988 | AUC = 0.8374

AMH IMPUTATION PERFORMANCE

- Among the four tested models, XGBoost gave the highest AMH-imputation R² (0.2267). DNN had the lowest imputation MAE (1.8279 ng/mL).
- XGBoost was selected for subsequent missingness experiments.

Model	R ²	MAE
Linear regression	0.2004	1.8720
XGBoost	0.2267	1.8331
LightGBM	0.2201	1.8378
DNN	0.2193	1.8279

INTERPRETABILITY (SHAP)

- AMH (observed or imputed) and age were the main prediction drivers. When AMH was omitted, age became the leading feature.
- FSH administration, infertility cause, hCG administration, and stimulation protocol provided additional information.

Feature hierarchy

- ✓ With AMH: AMH > age > FSH > infertility cause
- ✓ Without AMH: age > FSH > infertility cause > hCG / protocol

ROBUSTNESS TO MISSING AMH

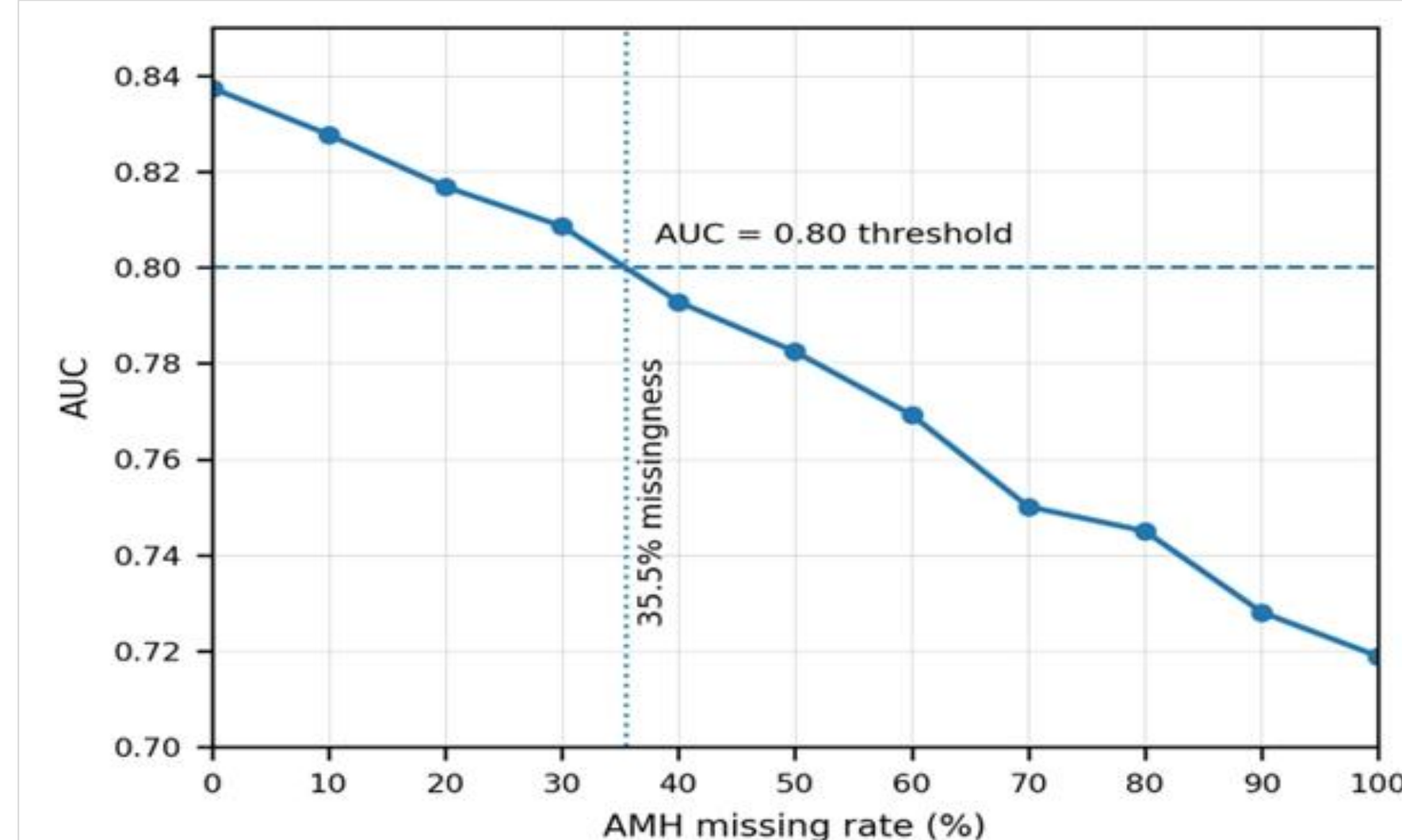


Fig. 3. Simulated AMH missingness versus XGBoost AUC. AUC crossed 0.80 at approximately 35.5% missingness.

AUC ≥ 0.80 through ~35.5% AMH missingness

SELECTED MISSINGNESS RESULTS

Missing AMH	AUC	R ²
0%	0.8374	0.4063
10%	0.8277	0.3786
20%	0.8168	0.3561
30%	0.8086	0.3357
40%	0.7928	0.3038
60%	0.7692	0.2572
100%	0.7189	0.1687

PRACTICAL DATA-QUALITY GUIDANCE

≤ 35% missing

AUC remained near or above 0.80; imputation may help preserve prediction.

35–60% missing

Performance declined; uncertainty and available covariates require attention.

> 60% missing

Prediction deteriorated more strongly; prioritize direct measurement or additional data.

CONCLUSIONS

- AI-based imputation preserved useful oocyte-response discrimination under moderate AMH missingness.
- Predictive performance should be evaluated downstream, not solely by biomarker-reconstruction accuracy.
- The ~35.5% threshold is study-specific—not a universal clinical cutoff.
- External validation and evaluation of non-random missingness remain necessary.

ACKNOWLEDGEMENT

This work was supported by the Electronics and Telecommunications Research Institute (ETRI) grant funded by the Korean government (26ZD1170), the National Research Foundation of Korea (RS-2026-25475659 and RS-2022-NR069006), and the Regional Innovation System & Education Program through Gyeongbuk RISE CENTER (2026-ANCHOR-15-108).

wh10287@naver.com(Hyo-Jin, Park)